

School's Performance Evaluation Based on Data Mining

Almuniri, Ismail; Said, Aiman Moyaid

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Almuniri, I., & Said, A. M. (2017). School's Performance Evaluation Based on Data Mining. *International Journal of Engineering and Information Systems*, 1(9), 56-62. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-55182-8>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/1.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/1.0>

School's Performance Evaluation Based on Data Mining

Ismail Almuniri, Aiman Moyaid Said

Department of Information Systems, University of Nizwa
07434064@uofn.edu.om, aiman.moyaid@unizwa.edu.om

Abstract: Educational sector is significant in any society as it plays a crucial role for developing the nation. Educational evaluation determines the effectiveness of the learning process and it interests all educational stakeholders. Currently, one of the challenges that the Ministry of Education in Oman is facing is to evaluate the performance of the school; this problem can be addressed through the use of modern technologies, such as the use of data mining. This study proposes to employ data mining to link between the different performance indicators with a goal of getting complete and effective evaluation for any school. Several classification techniques were exploited to analyze school performance. The results demonstrate a high accuracy level.

Keywords— Data Mining; Educational Data Mining; Classification

1. INTRODUCTION

The quality of education is the main goal of schools which it provides to its students. The best way to fulfill this level of quality in schools is through discovering knowledge. The knowledge discovery can be done by predicting the school performance. Performance evaluation serves the pedagogical field. It provides information to help students, teachers, administrators, and decision makers. Schools performance evaluation is now changing away from the traditional measurement way to modern evaluation techniques which employs various data which come from the educational system.

In Sultanate of Oman, Ministry of education has been implementing Educational portal since 2007. It can be considered as a learning management system (LMS). This portal provides many types of data in several sectors such as student, teachers and management. The ministry also aspires to the possibility of using the huge amount of data to benefit from the extraction of information and indicators that can serve the educational process and this is called Data Mining. The use of data mining methodologies is to analyze school's performance. Data mining provides many responsibilities that could be premonitory to check out school performance. Currently, schools have a massive set of data that has hidden valuable information. Through applying DM techniques we can get the useful information from this massive data set.

The current problem is that there is no mechanism for evaluating schools performance based on available data in the portal. Through this research, the authors will try to reach a mechanism to evaluate the performance of the school through data mining.

This paper is organized as follows. Section III provides a summary of some researches that have been done in EDM. Section IV provides the purpose and objective of the study. Section V discusses the method that has been used in this research. Section VI presents the experimental result and discusses the results. Finally Section VII provides the conclusion.

2. LITERATURE REVIEW

Data mining is defined by (Kamber, 2011) as "extracting" or "mining" knowledge from large amounts of data". From the point of view of the importance of education, and with the availability of vast amount of data that can be obtained from educational context; there is a need to implement the data mining technology to extract the information from the provided data. Many researches have been done in the field of Educational Data Mining (EDM), (Agarwal, Pandey, & Tiwari, 2012; Sachin & Vijay, 2011; Zimmermann, Brodersen, Heinemann, & Buhmann, 2015) (Member, n.d.) (Dutt, 2015).

(Galvan, 2016) studies the performance of the school and foregoes by his research about the control of administrative, management and operational procedures in public primary schools. The objective of his research was to develop a model through data mining and genetic algorithms. His research implementation was in a sample of public primary schools in Mexico.

S. Hari Ganesh and A. Joy Christy (Hari Ganesh & Joy Christy, 2015) defined data mining in their research paper as "a successful discipline in computer science that mines useful. There are increasing research interest in using data

mining in education. The Educational data mining (EDM) specialized in education sector, is an important field in data mining science. The most important researches dealing with the field of education is related to EDM (Baradwaj & Pal, 2012).

Predictive data mining tasks found a model from the accessible data set that is cooperative in predicting unknown or upcoming values of alternative data set of interest (Kamber, 2011).

After the rise of different methods which were proposed, applied and tested in the field of data mining, some researches argued that these generic methods or algorithms are not suitable to be applied to this upcoming field of study. It's suggested that the methods of educational data mining must be different from the traditional data mining ones because of multi-level hierarchy and non-independence in educational data. "Institutions are increasing being held accountable for student success since EDM emerged as a sub-discipline in DM there has been notable researches in student retention and attrition rates that have been conducted" (Dutt, 2015).

For the analysis of student's performance and to find the students that are performing consistently well, classification technique was used. The goal of applying data mining in this context is to reduce the student's drop out percentage and also raise school performance level. It is also used, by extracting the information from end of semester marks, to develop the student performance (Shaukat, Nawaz, Aslam, Zaheer, & Shaukat, 2016).

Parneet Kaur et al. (Kaur, Singh, & Josan, 2015) are identifying the slow learners and showing them by using three techniques which are Classification, Regression and Density Estimation. They used five Classification algorithms which are Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree to test and analyze the dataset by using cross validation.

Paulo Cortez and Alice Silva (Cortez & Gonçalves; 2008) approach student's achievements in secondary education using data mining by using past school grades. They collected data through school reports and questionnaires. They modeled two classes which are Mathematics and Protégés under five classification and regression tested through decision tree, Random forest, neural network and support vector machine. They reported this study by conducting the RMiner. The researchers came to the conclusion that student achievement is highly affected by previous performances.

3. PURPOSE AND OBJECTIVE

Educational sector is a very important part in our society and affects it. The Ministry of Education in Oman is currently facing difficulty in evaluating the performance of the school based on the current indicators which derive data from Oman's educational portal system. Also the ministry will be able to know the reasons for the weakness of some schools and how to solve them. The current system provides these indicators individually without any integration between them. This makes the school performance evaluation more difficult. The evaluation using DM can provide the ministry with many visions which serve to make the right decision as well as to get acquainted with high performance schools and to find out the reasons for this success.

With this challenge in the Ministry of Education and the availability of data from Oman's educational portal, this problem can be solved through the use of data mining methods. Data mining will enable us to link all indicators to get overall evaluation of the school. Among the areas of data mining, there is a specialized field to address the problems facing education field called Educational Data mining. Educational data mining is an upcoming domain which is productive and provides different types of information that can be estimated from learning goals. The purpose of this research is to evaluate the schools performance by taking different attributes from schools record such as learning achievement, Regularity and discipline students, Supervisory visits, Regularity and discipline staff, the development of school performance, Schools time tables and Vacations electronic system. Also, we analyze the impact of each attribute on the main attribute which is learning achievement. Because we need to predict the schools performance, we will use classification method in this research.

4. METHODOLOGY

Firstly, we will describe the dataset which is used in this research. Secondly, we will describe the attributes that we have used in this research. Finally, we will explain the classification techniques of the Data Mining that we have used.

Classification: It is a method of DM based on ML aiming which categorizes every item in a place of information into a place of a pre-defined program. Classification techniques make use of arithmetical and numerical methods that are; linear programming, neural network, support vector machines and decision trees. The aim of classification is to expect the target class for all the folders in the information (Shaukat et al., 2016).

4.1.1 One Rule: It is a simple algorithm classification that generates one rule for each predictor in the data, and then selects the rule with the smallest total error as its "one rule". We construct a frequency table for each predictor

against the target to create a rule for a predictor. It has been shown that OneR produces rules less accurate than state-of-the-art classification algorithms that are simple for humans to interpret. The working mechanism of oneR algorithm in each value of the predictor makes a rule is as following steps:

1. Count the number of class appearance
2. Find the highest class frequency
3. Make the rule assignment of that class to this value of the predictor
4. The total errors of the rules of each predictor are Calculated
5. Select the predictor with the smallest total error.

Accuracy within the OneR is practically more complex rules, have numerous implications for machine learning research and applications. From the usage of the 1R is to predict the accuracy within the produced rules by more sophisticated (ML) machine learning systems. The usage of the prediction can benchmark accuracy and it gives an estimate of how one learning system would compare with others. Also, it can be used to prove if learning is likely to reach predication in accuracy (Holte, 1993).

A OneR method is simple. It can create 1-level of decision tree expressed in the form of rules which test one attribute only. Also, it rises up with good rules for describing the structure of the data. (Ramaswami & Bhaskaran, 2009).

4.1.2 J48 grafts: This algorithm generates a grafted DT from a J48 tree. This technique of grafting adds nodes to an already existing decision tree to reduce prediction errors. It considers alternative classifications for identified regions of the instance space that are not occupied by training instances, or occupied only by misclassified training instances. Thus, the performance of the new test in the leaf is needed, creating new branches leading to new classifications. As a post- process, Grafting algorithm is for adding nodes to the tree. However, its main target is to increase the rightly classifying instances probability of falling outside the areas enclosed by the training data. Grafting can be practical to decision trees. Its goal is to decrease prediction error through reclassifying regions of the instance space where no training and misclassified data exist. Its purpose is to get the best matched cross with existing branches and leaf regions to create new leaves with other classifications. Here only branching that does not launch any classification errors in data already classified, though tree becomes more difficult. Thus, newly generated tree reduces errors instead of introducing them (Rajput & Arora, 2013).

4.1.3 J48: It is a version of an earlier ID3 algorithm developed by J. Ross Quinlan. The J48 algorithm in WEKA was used to test the data set and then after the result is shown for decision tree. The C4.5 algorithm is an open source Java implementation of J48. It uses measure called gain ratio to construct decision tree and generates non binary tree, the attribute with highest normalized gain ratio is taken as the root node and the dataset is split based on the root element values. The information gain is calculated for all the sub-nodes independently and the procedure is repeated until the prediction is finished. After the rising phase, Error-based pruning is performed. Both continuous and discrete attributes can be handled by J48 and also training data with missing attribute values and attributes with contradictory costs and afford an option to prune trees after creation. A number of parameters related to tree pruning in the J48 algorithm are provided and can make a remarkable difference in the results quality. J48 uses two post pruning methods, named sub tree rising and sub tree replacement (Rajput & Arora, 2013).

5. EEPREMENTAL AND SETUP RESULT

In this section we will talk about dataset, attributes and comparative results. Dataset contains required data needed for this research. The source of our data is The Educational Portal of Ministry of Education in Sultanate of Oman. In attributes, section we will clarify the mechanism of applying the selection methodology for each of the seven data indicators and how to derive the final values for each one. Finally, we will talk about comparative Results which contain explanation one technique which is classification and some of its algorithms we used in this research.

5.1 DATASET AND PREPROCESSING

The ministry is using different indicators for evaluating the performance of schools. The value of these indicators is store through Educational portal in the data repository. In this section we selected required data which needed for this research. The educational field is important because it has a massive data got from the educational system. In

sultanate Of Oman we have Educational Portal which is implemented since 2007. This portal provides many data regarding the students, teachers and leaning.

Currently, Educational Portal provides a data for seven indicators related to this research. These data need configuration to be valid in the analysis processes required in this research. The source of our data is The Educational Portal of Ministry of Education in Sultanate of Oman. Noting that the data were collected after the end of the first semester of the academic year 2016/2017 (for six months).A total of 261 schools were initially included in this research.

Here, explain the mechanism of applying the selection and transformation methodology for each of the seven indicators data and how to derive the final values for each indicator. The objective of this research is to analyze the school's performance by taking different attributes. The current performance indicators used by the ministry are shown in table 3.

Table 1: Attributes Definition

Attributes	Description
LA	Learning achievement, Describe the results of students in different assessment areas.
RDS	Regularity and discipline students, Describe the Regularity and discipline of the students.
SV	Supervisory visits, Define the Supervisory visits Statistically.
RDE	Regularity and discipline employee, Describe the Regularity and discipline of the employee.
SP	The development of school performance, Define the education, learning and school administration clauses.
STT	Schools time tables, mention the time table activation.
VES	Vacations electronic system, Shows extent of the school's activation of the electronic vacation system.

To prepare the data for analysis, several calculation were performs as explained below for each indicator.

5.1.1 *Learning achievement*: This performance indicator shows the result of all students in all courses in each semester. The indicator shows the general ratio of the school in each course by showing the percentage of each course in each level. The five levels are A - B - C – D and F. The data in this indicator need to be summarized at the level of school. The last form of data will be by calculating the total percentage in each school level.

5.1.2 *Regularity and discipline students*: This performance indicator shows the regularity and discipline of the students in each semester. It provides much information related to the students regularity. Currently, it shows the number of cases in each category. The categories are Total Absence, Unexcused absence, Excused absence, Total delay, written advice, Alert the student, Warning the student with a pledge from the parent, Study student's case, raising the case of the student to the committee, temporary separation and commitment to the student and parent, decreasing continuous ratings marks, Deprivation from the semester examinations, Transfer student disciplinary, Deposition in the same academic year and Final suspend.

5.1.3 *Supervisory visits*: This performance indicator shows the number of all visit forms in all schools and regions in each semester. It, also, shows the number of the visits at the level of Region, school and subject. In addition, it shows the order of the schools in each region based on the number of the visits.

5.1.4 *Regularity and discipline staff*: This performance indicator extrapolates the loss of working days in accordance with absenteeism with or without the excuse and loss of learning time through the extrapolation of two basic elements, namely late attendance in the morning and early leave.

5.1.5 *The development of school performance*: This performance indicator shows the average of the three clauses which are education, learning and school administration. It, also, presents the evaluation according to the general criterion and the type of form. The three fields are a collection of forms to be filled in through which each of the three items is evaluated separately. This indicator is the ranking of schools according to the general average at the region level.

5.1.6 *Schools time tables*: This performance indicator shows the level of the schools' time table activation in the educational portal. It shows according to the approved study plan and the manual of the functions of the school posts and the approved work load.

5.1.7 *Vacations electronic system*: The performance indicator shows the extent of the school's activation of the electronic vacation system. It provides many charts and numbers according to levels.

The Next step is Data cleaning. This step permits transforming the available data into an appropriate format to be ready to use. Later, data mining algorithm can be used to solve a specific problem.

5.2 COMPARATIVE RESULTS

The primary objective of the study is to evaluate the performance of the schools effectively. For this purpose, it is explained that one technique which is classification and some of its algorithms we used in this research. The following table shows confusion matrix measurement and the description of each measure.

Table 2: Confusion Matrix Measurement and Description

Measurements	Description
Correctly classified	Mention the percentage of the test correctness.
Inaccurately classified	Mention the percentage of the test incorrectness.
TP Rate	The average of the true positive.
FP Rate	The average of the False positive.
Precision	Ratio of TP divided by the sum of TP and FP.
Accuracy	Ratio of sum TP and TN divided by the total number of instance.
Error Rate	One Subtract Accuracy.

It implemented a cross-validation to build a generalize model also to avoid overfill problem. Furthermore, I have used ten folders in a cross-validation.

The Fig 1 and table 2 show the results of three algorithms which are OneR, J48 grafts and J48. They provide the confusion matrix result in the three cases. Also show correctly classified, inaccurately classified, true positive, false positive, The Precision, Accuracy and the error rate. By observing the figure above, it shows that the oneR algorithm has the best result and highest accuracy. Whereas the correctly classified is the highest and equal 86.21% and inaccurately classified is 13.79%. The correctly classified of the J48 grafts algorithm is 85.06% and inaccurately classified equal 14.94%. The correctly classified of the J48 algorithm is the less one compare to the other algorithms and equal 84.67% and inaccurately classified is 15.33%

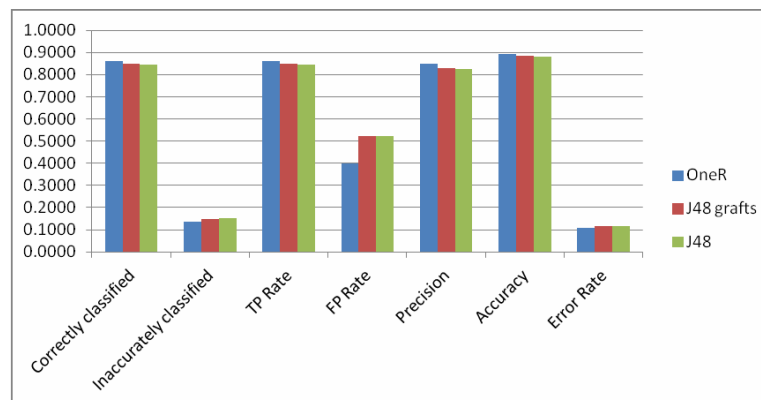


Fig 1: Comparison graph of Classification Techniques

Table 3: Comparison Table Of Classification Techniques

Algorithm	Correctly classified	Inaccurately classified	TP Rate	FP Rate	Precision	Accuracy	Error Rate
OneR	86.21%	13.79%	0.862	0.398	0.851	0.8934	0.1066
J48 grafts	85.06%	14.94%	0.851	0.523	0.828	0.8851	0.1149
J48	84.67%	15.33%	0.847	0.524	0.824	0.8824	0.1176

6. CONCLUSION:

In this paper we explained the classification technique of the Data Mining and some algorithms of these techniques. In Classification, we have used the three techniques of the classification that are OneR, J48 grafts and J48 to classify our training data set. We have mentioned the result of every technique separately. The results of these techniques are in the form of figures no4, 5 and 6 and table no. 2.

REFERENCES:

- [1] Abu Amuna, Y., Al Shobaki, M., Naser, S. S. A., & Badwan, J. (2017). Understanding Critical Variables For Customer Relationship Management In Higher Education Institution From Employee Perspective. *International Journal of Information Technology and Electrical Engineering*, 6(1), 10-16.
- [2] Abu Naser, S. S., & Al Shobaki, M. J. (2016). Computerized Management Information Systems Resources and their Relationship to the Development of Performance in the Electricity Distribution Company in Gaza. *EUROPEAN ACADEMIC RESEARCH*, 6(8), 6969-7002.
- [3] Abu Naser, S. S., & Al Shobaki, M. J. (2016). The Impact of Management Requirements and Operations of Computerized Management Information Systems to Improve Performance (Practical Study on the employees of the company of Gaza Electricity Distribution). Paper presented at the First Scientific Conference for Community Development.
- [4] Abu Naser, S. S., & Sulisel, O. (2000). The effect of using computer aided instruction on performance of 10th grade biology in Gaza. *Journal of the College of Education*, 4, 9-37.
- [5] Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data Mining in Education : Data Classification and Decision Tree Approach. *International Journal of E-Education, E- Business, E-Management and E-Learning*, 2(2), 140–145. Retrieved from papers3://publication/uuid/88A64A7E-BAB5-4E90-8192-2C3F1886CB9A
- [6] Al Shobaki, M. J., & Abu Naser, S. S. (2016). Performance development and its relationship to demographic variables among users of computerized management information systems in Gaza electricity Distribution Company. *International Journal of Humanities and Social Science Research*, 2(10), 21-30.
- [7] Al Shobaki, M. J., & Abu Naser, S. S. (2016). The Dimensions Of Organizational Excellence In The Palestinian Higher Education Institutions From The Perspective Of The Students. *GLOBAL JOURNAL OF MULTIDISCIPLINARY STUDIES*, 5(11), 66-100.
- [8] Al Shobaki, M. J., & Abu Naser, S. S. (2017). The Role of the Practice of Excellence Strategies in Education to Achieve Sustainable Competitive Advantage to Institutions of Higher Education-Faculty of Engineering and Information Technology at Al-Azhar University in Gaza a Model. *International Journal of Digital Publication Technology*, 1(2), 135-157.
- [9] Baradwaj, B., & Pal, S. (2012). Mining educational data to analyze student's performance.
- [10] Cortez, P. S., & Gonçalves, A. M. (2008). Using data mining to predict secondary school student performance. *Proceedings of 5th Annual Future Business Technology Conference*, 2003(2000), 5–12. Retrieved from <http://hdl.handle.net/1822/8024>
- [11] Dutt, A. (2015). Clustering Algorithms Applied in Educational Data Mining. *International Journal of Information and Electronics Engineering (IJIEE)*, 5(2), 112–116. <https://doi.org/10.7763/IJIEE.2015.V5.513>
- [12] Galvan, P. (2016). Educational evaluation and prediction of school performance through data mining and genetic algorithms. *2016 Future Technologies Conference (FTC)*, (December), 245–249. <https://doi.org/10.1109/FTC.2016.7821617>
- [13] Hari Ganesh, S., & Joy Christy, A. (2015). Applications of Educational Data Mining: A survey.
- [14] Holte, R. C. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Machine Learning*, 11(1), 63–91. <https://doi.org/10.1023/A:1022631118932>
- [15] ICIECS 2015 - 2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems. <https://doi.org/10.1109/ICIECS.2015.7192945>
- [16] *International Journal of Advanced Computer Science and Applications*, 2(6), 63–69. <https://doi.org/vol.2,No.6>
- [17] Kamber, J. and M. (2011). *Data Mining: Concepts and Techniques*. Elsevier (Vol. 12). <https://doi.org/10.1007/978-3-642-19721-5>
- [18] Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 57, 500–508. <https://doi.org/10.1016/j.procs.2015.07.372>
- [19] Member, S. (n.d.). For Peeevr Riew For Peeevw, XX(X), 1–19.
- [20] Rajput, S., & Arora, A. (2013). Designing Spam Model-Classification Analysis using Decision Trees. *International Journal of Computer Applications*, 75(10), 975–8887.
- [21] Ramaswami, M., & Bhaskaran, R. (2009). A Study on Feature Selection Techniques in Educational Data Mining. *Journal of Computing*, 1(1), 7–11. Retrieved from <http://arxiv.org/abs/0912.3924>
- [22] Sachin, R. B., & Vijay, M. S. (2011). A survey and future vision of data mining in educational field. *Proceedings - 2012 2nd International Conference on Advanced Computing and Communication Technologies, ACCT 2012*, 96–100. <https://doi.org/10.1109/ACCT.2012.14>

- [23] Shaukat, K., Nawaz, I., Aslam, S., Zaheer, S., & Shaukat, U. (2016). Student's performance in the context of data mining. 2016 19th International Multi-Topic Conference (INMIC), 1–8. <https://doi.org/10.1109/INMIC.2016.7840072>
- [24] Zimmermann, J., Brodersen, K. H., Heinemann, H. R., & Buhmann, J. M. (2015). A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance. JEDM - Journal of Educational Data Mining, 7(3), 151–176. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM070/pdf_